INDIAN INSTITUTE OF SCIENCE

STOCHASTIC HYDROLOGY

Lecture -33

Course Instructor : Prof. P. P. MUJUMDAR
Department of Civil Engg., IISc.

Summary of the previous lecture

- Regression on Principal components

$$x_{ij} = \frac{\left(X_{ij} - \bar{x}_j\right)}{s_j} \qquad y_i = Y_i - \bar{y}$$

$$Z = X\,A$$

$$Y = Z\,\mathrm{B}$$

MULTIVARIATE STOCHASTIC MODELS

Multivariate Stochastic models

- Stochastic models discussed for single site in relation with the auto correlations and auto covariance.
  - Thomas Fiering models
    - Stationary and non-stationary models
  - ARMA models
    - Box Jenkins models

$\{x_t\}$

x_{t+1}

Multivariate Stochastic models

First order Markov process:

Random component

$$X_{t+1} = \underbrace{\mu_x + \rho_1 (X_t - \mu_x)}_{\text{Deterministic component}} + \varepsilon_{t+1}$$

ε ~ Mean 0 and variance σ_ε^2

$$X_{t+1} = \mu_x + \rho_1 \left(X_t - \mu_x \right) + u_{t+1} \sigma_x \sqrt{1 - \rho_1^2}$$

Multivariate Stochastic models

First order Markov model with non-stationarity, for stream flow generation:

$$X_{i,j+1} = \mu_{j+1} + \rho_j \frac{\sigma_{j+1}}{\sigma_j}\left(X_{ij} - \mu_j\right) + t_{i,j+1}\sigma_{j+1}\sqrt{1-\rho_j^2}$$

ρ_j is serial correlation between flows of j^{th} month and $j+1^{th}$ month.

$t_{i,\, j+1} \sim N(0, 1)$

ARIMA Models

ARMA (p, q)

Residuals of order 'q'

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q} + e_t$$

AR of order 'p'

$\{e_t\}$ is the residual series

Assumptions : $\{e_t\}$ has zero mean with uncorrelated terms

$$\hat{X}_{t+1} = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=1}^{q} \theta_j e_{t-j}$$

Multivariate Stochastic models

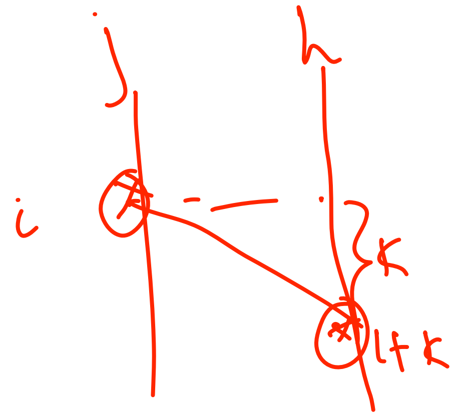
- Data generation (or forecasting) on a random variate depending on two or more sites is usually required.
  - For example, in the design of a reservoir, the flow from all the streams fed to the reservoir must be considered.
- If the time series for the random variables are independent, then the generation techniques for single site can be used.
- When they are not independent, it is important to consider the simultaneous behavior of the random variables.

Multivariate Stochastic models

- Correlation of a random variable between two sites is cross-correlation.
- Lag zero cross-correlation is the correlation of a random variable at two points in the same time period.
- Lag k cross-correlation, $r_{j,h}(k)$ is the correlation between random variable at site j with the random variable at site h, with lag time k.

Multivariate Stochastic models

$$r_{j,h}\left(k\right)=\frac{\sum_{i=1}^{n}\left(x_{j,i}-\overline{x}_{j}\right)\left(x_{h,i+k}-\overline{x}_{h}\right)}{\left(n-k\right)s_{j}s_{h}}$$



where

n is the total number of pairs of observations on X_j and X_k,

$x_{j,i}$ is the ith observation on X_j

$\overline{x}_j, s_j$ are the mean and standard deviation of the observations on X_j

Example – 1

Obtain the lag one cross correlation of annual rainfall data in mm at two sites A and B.

Year	1	2	3	4	5	6	7	8	9	10
Annual rainfall at site A (mm)	5496	7797	7392	7061	6564	5919	5053	3951	4280	5910
Annual rainfall at site B (mm)	5713	6934	6275	6641	6675	5605	5144	5116	4722	6869

Year	11	12	13	14	15	16	17	18	19
Annual rainfall at site A (mm)	5145	6384	5679	6021	6733	8151	4151	4200	6704
Annual rainfall at site B (mm)	5226	7313	6068	5876	6044	8384	5149	5359	6197

Example – 1 (Contd.)

Site	A	B
Mean	5926	6069
Std.dev.	1250.1	914.9

lag one cross correlation of sites A and B is given by

$$r_{A,B}\left(1\right) = \frac{\sum_{i=1}^{n}\left(x_{A,i} - \overline{x}_A\right)\left(x_{B,i+1} - \overline{x}_B\right)}{\left(n-1\right)s_A s_B}$$

S.No. (i)	Annual rainfall at A	Annual rainfall at B	$\left(x_{A,i}-\overline{x}_A\right)$	$\left(x_{B,i+1}-\overline{x}_B\right)$	$\left(x_{A,i}-\overline{x}_A\right)\left(x_{B,i+1}-\overline{x}_B\right)$
1	5496	5713	-430		
2	7797	6934	1871	865	-371836
3	7392	6275	1466	206	385557
4	7061	6641	1135	572	838719.5
5	6564	6675	638	606	687965.4
6	5919	5605	-7	-464	-296072
7	5053	5144	-873	-925	6328.587
8	3951	5116	-1975	-953	831772.6
9	4280	4722	-1646	-1347	2660008
10	5910	6869	-16	800	-1316760
11	5145	5226	-781	-843	13354.06
12	6384	7313	458	1244	-971409
13	5679	6068	-247	-0.95	-434.044
14	6021	5876	95	-193	47627.53
15	6733	6044	807	-25	-2373.94
16	8151	8384	2225	2315	1868613
17	4151	5149	-1775	-920	-2047028
18	4200	5359	-1726	-710	1260044
19	6704	6197		128	-220999
				Σ	3373079

Example – 1 (Contd.)

$$\sum_{i=1}^{n}\left(x_{A,i}-\overline{x}_A\right)\left(x_{B,i+1}-\overline{x}_B\right)=3373079$$

$$r_{A,B}\left(1\right)=\frac{\sum_{i=1}^{n}\left(x_{A,i}-\overline{x}_A\right)\left(x_{B,i+1}-\overline{x}_B\right)}{\left(n-1\right)s_A s_B}$$

$$=\frac{3373079}{\left(19-1\right)\times 1250.1\times 914.9}$$

$$=0.164$$

Multivariate Stochastic models

Multisite Markov model (Two sites):

- Model preserves mean, variance, skewness, lag one serial correlation and lag zero cross-correlation (Haan 1977).
- One site is to be selected as key site.
- Selection may be based on the length of the data and the quality of the record.
- Consider j as the key site and h as the subordinate site to key site j.
- A sequence of observations is generated for site j using single site generation technique.

Ref.: Haan, C.T. (1977) Statistical methods in Hydrology, Iowa State University Press

Multivariate Stochastic models

- A cross-correlation model is used to generate values of site h based on generated values at site j.

$$X_{h,t} = \bar{x}_h + r_{j,h}(0)\frac{s_h}{s_j}\left(X_{j,t} - \bar{x}_j\right) + u_t s_h \sqrt{1 - r_{j,h}^2(0)}$$

j and *h* refer to two sites, in this model

First order Markov model with non-stationarity (single site)

$$X_{i,j+1} = \mu_{j+1} + \rho_j \frac{\sigma_{j+1}}{\sigma_j}\left(X_{ij} - \mu_j\right) + t_{i,j+1}\sigma_{j+1}\sqrt{1 - \rho_j^2}$$

i is year *j* is month in this model

Multivariate Stochastic models

where

u_t is a standardized random variate adjusted to incorporate the serial correlation at site h.

$$u_t = \zeta \frac{\left(X_{h,t-1} - \overline{x}_h\right)}{s_h} + t_t\sqrt{1-\zeta^2}$$

t_t is a standardized random variate

Serial Correlation

$$\zeta = \frac{r_h\left(1\right) - r_j\left(1\right) r_{j,h}^2\left(0\right)}{\sqrt{1 - r_{j,h}^2\left(0\right)}}$$

Multivariate Stochastic models

Multisite Markov model:

- Multisite generation requires simultaneous generation of data at several sites while preserving the correlation between the data at various sites.
- Consider $x_{j,t}$

$$x_{j,t} = \frac{\left(x_{j,t} - \overline{x}_j\right)}{s_j}$$

Multivariate Stochastic models

$$X_{t+1} = \mu_x + \rho_1 \left(X_t - \mu_x \right) + u_{t+1} \sigma_x \sqrt{1 - \rho_1^2}$$

- The first order Markov model for site h is

$$x_{h,t+1} = \rho_h \left(1\right) x_{h,t} + \varepsilon_{h,t+1} \sqrt{1 - \rho_h^2 \left(1\right)}$$

μ= 0 and σ =1 because it is standardized data

- The first order Markov model for site j is

$$x_{j,t+1} = \rho_j \left(1\right) x_{j,t} + \varepsilon_{j,t+1} \sqrt{1 - \rho_j^2 \left(1\right)}$$

- The equations are written in matrix form

Multivariate Stochastic models

$$X_{t+1} = EX_t + G\varepsilon$$

where

X_t is a p x 1 vector of standardized values of the variable generated at time t,

E is a p x p diagonal matrix whose jth diagonal element is $\rho_j(1)$,

G is a p x p diagonal matrix whose jth diagonal element is $\sqrt{1-\rho_j^2(1)}$

ε is a p x 1 vector of random variates

Multivariate Stochastic models

- ε is defined to preserve the first order serial correlation (auto correlation) of the x_j 's and the lag zero cross-correlation between x_j and x_h .
- ε is made of elements that are $\varepsilon_{j,t+1}$; each $\varepsilon_{j,t+1}$ is independent of $x_{j,t}$; ε_j is N(0,1)
- The cross correlation between ε_j and ε_h is $\rho^*_{j,h}(0)$,

$$\rho^*_{j,h}(0) = \frac{\left\{1-\rho_j(1)\rho_h(1)\right\}\rho_{j,h}(0)}{\sqrt{\left\{1-\rho_j^2(1)\right\}\left\{1-\rho_h^2(1)\right\}}}$$

- $\rho^*_{j,h}(0)$ reproduces the desired $\rho_{j,h}(0)$, which is the lag zero cross correlation between x_j and x_h .

Multivariate Stochastic models

$$\varepsilon = AD_{\lambda}^{1/2}e$$

where

$D_{\lambda}^{1/2}$ is a p x p diagonal matrix whose j^{th} diagonal element is the square root of the j^{th} largest eigenvalue of the p x p correlation matrix whose elements are $\rho *_{j,h}(0)$

A is a p x p matrix consisting of eigenvectors of correlation matrix,

e is p x 1 vector of independent observations from N (0,1)

Multivariate Stochastic models

- Matalas (1967) has given a multisite normal generation model that preserves the means, variances, lag one serial correlation, lag one cross-correlations and lag zero cross-correlations.

$$X_{t+1} = AX_t + B\varepsilon_{t+1}$$

where

X_t and X_{t+1} are p x 1 vectors representing standardized data corresponding to p sites at time steps t and $t+1$ resp.

Ref.: Matalas, N.C. (1967) Mathematical assessment of synthetic hydrology, Water Resources Research 3(4):937-945

Multivariate Stochastic models

ε_{t+1} is a form of N(0,1) with ε_{t+1} independent of X_t.

A and B are coefficient matrices of size p x p.

$$X_{t+1} = \begin{bmatrix} x(1,t+1) \\ x(2,t+1) \\ . \\ . \\ x(i,t+1) \\ . \\ . \\ x(p,t+1) \end{bmatrix} \qquad X_t = \begin{bmatrix} x(1,t) \\ x(2,t) \\ . \\ . \\ x(i,t) \\ . \\ . \\ x(p,t) \end{bmatrix} \qquad \varepsilon_{t+1} = \begin{bmatrix} \varepsilon(1,t+1) \\ \varepsilon(2,t+1) \\ . \\ . \\ \varepsilon(i,t+1) \\ . \\ . \\ \varepsilon(p,t+1) \end{bmatrix}$$

Multivariate Stochastic models

- The equation form is

$$x_{i,t+1} = \sum_{j=1}^{p} a_{i,j} x(i,t) + \sum_{j=1}^{i} b_{i,j} \varepsilon(i,t+1)$$

where

$a_{i,j}$ and $b_{i,j}$ denote the (i , j)th elements of the matrices A and B.

B is assumed to be lower triangular matrix.

Multivariate Stochastic models

Coefficient matrices A and B:

- The expectation of $X_t X_t'$ is denoted by M_0

$$M_0 = E\left[X_t X_t'\right]$$

If $m_0(i\ ,j)$ is a element of M_0 matrix in the i^{th} row and j^{th} column,

$$m_0(i,j) = \frac{1}{n}\left[x(i,1)x(j,1) + x(i,2)x(j,2) + ...x(i,n)x(j,n)\right]$$

$$m_0(i,j) = \frac{1}{n}\sum_{t=1}^{n} x(i,t)x(j,t)$$

Multivariate Stochastic models

$$m_0\left(i,j\right)=\frac{1}{n}\sum_{t=1}^{n}\left(\frac{X_{i,t}-\overline{x}_i}{s_i}\right)\left(\frac{X_{j,t}-\overline{x}_j}{s_j}\right)$$

i.e., $m_0(i\ ,j)$ is correlation coefficient between the data at sites i and j at time t.

Therefore M_0 is the cross-covariance matrix of lag zero

Multivariate Stochastic models

- The expectation of $X_t X_{t-1}'$ is denoted by M_1

$$M_1 = E\left[X_t X_{t-1}^{'}\right]$$

If $m_1(i\ ,j)$ is a element of M_1 matrix in the i^{th} row and j^{th} column,

$$m_1(i,j) = \frac{1}{n}\left[x(i,1)x(j,0) + x(i,2)x(j,1) + \ldots\ldots x(i,n)x(j,n-1)\right]$$

$$m_1(i,j) = \frac{1}{n-1}\sum_{t=2}^{n} x(i,t)x(j,t-1)$$

Multivariate Stochastic models

$$m_1(i,j) = \frac{1}{n-1}\sum_{t=2}^{n}\left(\frac{X_{i,t} - \overline{x}_i}{s_i}\right)\left(\frac{X_{j,t-1} - \overline{x}_j}{s_j}\right)$$

i.e., $m_1(i\ ,j)$ represents lag one cross correlation between the data at sites i and j .

Therefore M_1 is the cross-covariance matrix of lag one

Multivariate Stochastic models

Considering the model,

$$X_{t+1} = AX_t + B\varepsilon_{t+1}$$

Post multiplying with X_t' on both sides and taking the expectation,.

$$E\left[X_{t+1}X_t'\right] = AE\left[X_tX_t'\right] + B\underbrace{E\left[\varepsilon_{t+1}X_t'\right]}_{0}$$

$$M_1 = AM_0 + 0$$

$$A = M_1M_0^{-1}$$

Multivariate Stochastic models

Post multiplying with $X_{t+1}{}'$ on both sides and taking the expectation,.

$$X_{t+1} = AX_t + B\varepsilon_{t+1}$$

$$\underbrace{E\left[X_{t+1}X_{t+1}^{'}\right]}_{M_0} = AE\left[X_t X_{t+1}^{'}\right] + BE\left[\varepsilon_{t+1}X_{t+1}^{'}\right]$$

Multivariate Stochastic models

$$M_1 = E\left[X_t X_{t-1}^{'}\right]$$

$$M_1^{'} = \left\{E\left[X_t X_{t-1}^{'}\right]\right\}^{'}$$

$$= E\left[\left\{X_t X_{t-1}^{'}\right\}^{'}\right]$$

$$= E\left[X_{t-1} X_t^{'}\right]$$

or

$$M_1^{'} = E\left[X_t X_{t+1}^{'}\right]$$

Multivariate Stochastic models

$$\varepsilon_{t+1} X_{t+1}^{'} = \varepsilon_{t+1} \left\{ A X_t + B \varepsilon_{t+1} \right\}^{'}$$

$$= \varepsilon_{t+1} X_t^{'} A^{'} + \varepsilon_{t+1} \varepsilon_{t+1}^{'} B^{'}$$

Taking expectation on both sides,

$$E\left[\varepsilon_{t+1} X_{t+1}^{'}\right] = E\left[\varepsilon_{t+1} X_t^{'} A^{'} + \varepsilon_{t+1} \varepsilon_{t+1}^{'} B^{'}\right]$$

$$= E\left[\varepsilon_{t+1} X_t^{'} A^{'}\right] + E\left[\varepsilon_{t+1} \varepsilon_{t+1}^{'} B^{'}\right]$$

$$= 0 + I B^{'}$$

$$= B^{'}$$

Multivariate Stochastic models

Substituting in the equation,

$$E\left[X_{t+1}X_{t+1}^{'}\right] = AE\left[X_{t}X_{t+1}^{'}\right] + BE\left[\varepsilon_{t+1}X_{t+1}^{'}\right]$$

$$M_0 = AM_1^{'} + BB^{'}$$

$$A = M_1 M_0^{-1}$$

$$M_0 = M_1 M_0^{-1} M_1^{'} + BB^{'}$$

$$BB^{'} = M_0 - M_1 M_0^{-1} M_1^{'}$$

If $C = BB'$

$$C = M_0 - M_1 M_0^{-1} M_1^{'}$$

Multivariate Stochastic models

- B does not have a unique solution.
- One method is to assume B is a lower triangular matrix.

$$BB' = \begin{bmatrix} b(1,1) & 0 & 0 & . & . & 0 \\ b(2,1) & b(2,2) & 0 & . & . & 0 \\ . & . & & & & \\ . & . & & & & \\ b(p,1) & b(p,2) & . & . & . & b(p,p) \end{bmatrix} \begin{bmatrix} b(1,1) & b(2,1) & . & . & . & b(p,1) \\ 0 & b(2,2) & . & . & . & b(p,2) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & 0 & . & . & . & b(p,p) \end{bmatrix}$$

$$C = \begin{bmatrix} c(1,1) & c(1,2) & c(1,3) & . & . & c(1,p) \\ b(2,1) & b(2,2) & b(2,3) & . & . & b(2,p) \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ b(p,1) & b(p,2) & . & . & . & b(p,p) \end{bmatrix}$$

Multivariate Stochastic models

- The diagonal elements of the B matrix are obtained as,

$$b(1,1) = c(1,1)^{1/2}$$

$$b(2,2) = \left\{c(2,2) - b^2(2,1)\right\}^{1/2}$$

.

.

.

$$b(k,k) = \left\{c(k,k) - b^2(k,k-1) - b^2(k,k-2) - \ldots - b^2(k,1)\right\}^{1/2}$$

Multivariate Stochastic models

- The elements in the k^{th} row are obtained as,

$$b(k,1) = \frac{c(k,1)}{b(1,1)}$$

$$b(k,2) = \frac{c(k,2) - b(2,1)b(k,1)}{b(2,2)}$$

.

.

$$b(k,j) = \frac{c(k,j) - b(j,1)b(k,1) - b(j,2)b(k,2)......b(j,j-1)b(k,j-1)}{b(j,j)}$$

Multivariate Stochastic models

- If the model is to fit the data, the matrices M_0 and BB' should be positive definite.
- This condition is used to check the inconsistency in the data.
- Assumption is that the model is multivariate normal.